

THE INVENTION CLAIMED IS:

1. A method for categorization of an item comprising:
providing a plurality of categories organized in a hierarchy of categories;
providing a plurality of categorizers corresponding to the plurality of categories;
5 featurizing the item to create a list of item features;
using the list of item features in a categorizer system including the plurality of
categorizers for determining a plurality of levels of goodness;
using one of the plurality of levels of goodness for invoking an additional categorizer
of the plurality of categorizers as required;
10 categorizing the item in the categorizer system in the plurality of categories based on
the respective plurality of levels of goodness; and
returning the item categorized.
2. The method as claimed in claim 1 wherein:
using the list of item features determines the plurality of levels of goodness using a
15 process to quantify the plurality of levels of goodness, to prioritize the
plurality of levels of goodness, and to resolve two levels of goodness into a
third level of goodness.
3. The method as claimed in claim 1 including:
using a categorizer system knowledge base for determining the level of goodness for a
20 category with the list of item features.
4. The method as claimed in claim 1 including:
listing the plurality of categories and the respective levels of goodness on a list; and
categorizing from the list.
5. The method as claimed in claim 1 wherein:
25 returning one category for the item among the plurality of categories selected from a
group consisting of the one category with the best level of goodness for all the
plurality of categories and with the best level of goodness for which
determining is completed where all of the plurality of categories are not
compared.
- 30 6. The method as claimed in claim 1 wherein:
returning a plurality of categories for the item among the plurality of categories
returns a plurality of categories selected from a group consisting of categories
up to a fixed number of the plurality of categories, categories having more

than a fixed level of goodness, categories fulfilling a user specified preference, categories not from a categorizer, and categories which are a combination thereof.

7. The method as claimed in claim 1 wherein:

returning the category for a plurality of items establishes a categorizer system knowledge base for a topic hierarchy.

8. The method as claimed in claim 1 including:

listing a plurality of labels for each of the plurality of categories; and

training a categorizer system trainer using a plurality of items having known categories and the plurality of labels to provide a categorizer system knowledge base.

9. The method as claimed in claim 1 including:

providing a categorizer system knowledge base;

using a plurality of items with known categories to learn knowledge in the categorizer system knowledge base.

10. The method as claimed in claim 1 including:

providing a categorizer system knowledge base;

providing a plurality of categorizers, each using knowledge in a categorizer system knowledge base and the list of item features to compute a degree of goodness for a plurality of categories, independent of other categorizers, each using a subset of item features to compute a degree of goodness for a plurality of categories, independent of other categorizers, and each subset independent of subsets used by other categorizers; and

providing a mechanism to resolve the levels of goodness for a plurality of categories resulting from multiple categorizers into a combined level of goodness for a plurality of categories.

11. A method for categorization of an item comprising:

providing a plurality of categories organized in a hierarchy of categories and having respective lists of category features using a categorizer system knowledge base for determining the lists of category features;

providing a plurality of categorizers corresponding to one of the plurality of categories;

featurizing the item to create a list of item features;

using the list of item features in a categorizer system including the plurality of categorizers with the lists of category features to respectively determine a plurality of levels of goodness, the plurality of levels of goodness determined using a process to quantify the plurality of levels of goodness, to prioritize the plurality of levels of goodness, and to resolve two levels of goodness into a third level of goodness;

using one of the plurality of levels of goodness for invoking an additional categorizer of the plurality of categorizers as required;

categorizing the item in the categorizer system in the plurality of categories based on the respective plurality of levels of goodness;

listing the plurality of categories and the respective levels of goodness on a list; and returning a category for the item from the list.

12. A method for categorization of a document comprising:

providing a plurality of categories organized in a hierarchy of categories;

providing a plurality of categorizers corresponding to the plurality of categories;

featurizing the document to create a list of document features;

using the list of document features in a categorizer system including the plurality of categorizers for determining a plurality of levels of goodness;

using one of the plurality of levels of goodness for invoking an additional categorizer of the plurality of categorizers as required;

categorizing the document in categorizer system in the plurality of categories based on the respective plurality of levels of goodness; and

returning a category for the document.

13. The method as claimed in claim 12 wherein:

determining the plurality of levels of goodness includes using a process selected from a group consisting of Naïve Bayes, quantitative decision-tree classifiers such as C4.5, Bayesian networks, rule-based multi-class classifiers that output a degree of goodness, conditional probability statements, simple heuristics, and a combination thereof.

14. The method as claimed in claim 12 including:

using a categorizer system knowledge base for determining the level of goodness for a category with the list of document features.

15. The method as claimed in claim 12 including:
listing the plurality of categories as the document is compared and the respective
levels of goodness on a list; and
categorizing from the list.

5 16. The method as claimed in claim 12 wherein:
returning one category for the document among the plurality of categories selected
from a group consisting of the one category with the best level of goodness for
all the plurality of categories and with the best level of goodness for which
determining is completed where all of the plurality of categories are not
10 compared.

17. The method as claimed in claim 12 wherein:
returning a plurality of categories for the document among the plurality of categories
returns a plurality of categories selected from a group consisting of categories
up to a fixed number of the plurality of categories, categories having more
15 than a fixed level of goodness, categories fulfilling a user specified preference,
categories not from a categorizer, and categories which are a combination
thereof.

18. The method as claimed in claim 12 wherein:
returning the category for a plurality of documents establishes a categorizer system
20 knowledge base for a topic hierarchy.

19. The method as claimed in claim 12 including:
listing a plurality of labels for each of the plurality of categories; and
training a categorizer system trainer using a plurality of documents having known
categories and the plurality of labels to provide a categorizer system
25 knowledge base.

20. The method as claimed in claim 12 including:
providing a categorizer system knowledge base;
using a plurality of documents with known categories to learn knowledge in the
categorizer system knowledge base.

30 21. The method as claimed in claim 12 including:
providing a categorizer system knowledge base;
providing a plurality of categorizers, each using knowledge in a categorizer system
knowledge base and the list of document features to compute a degree of

goodness for a plurality of categories, independent of other categorizers, each using a subset of document features to compute a degree of goodness for a plurality of categories, independent of other categorizers, and each subset independent of subsets used by other categorizers; and

5 providing a mechanism to resolve the levels of goodness for a plurality of categories resulting from multiple categorizers into a combined level of goodness for a plurality of categories.

22. A method for categorization of a document comprising:

10 providing a plurality of categories organized in a hierarchy of categories and having respective lists of category features using a categorizer system knowledge base resulting from determining a plurality of documents for determining the lists of category features;

providing a plurality of categorizers corresponding to the plurality of categories;

featurizing the document to create a list of document features;

15 using the list of document features in a categorizer system including the plurality of categorizers with the lists of category features to respectively determine a plurality of levels of goodness;

using one of the plurality of levels of goodness for invoking an additional categorizer of the plurality of categorizers as required;

20 categorizing the document in categorizer system including the plurality of categorizers in the plurality of categories based on the respective plurality of levels of goodness from the list;

listing the plurality of categories as the document is compared and the respective levels of goodness on a list; and

25 returning a category for the document from the list.

23. A system for categorization of an item comprising:

a plurality of categories organized in a hierarchy of categories;

a plurality of categorizers corresponding to the plurality of categories;

a featurizer for featurizing the item to create a list of item features;

30 a categorizer system including the plurality of categorizers using the list of item features in for determining a plurality of levels of goodness, the categorizer system for categorizing the item in the plurality of categories based on the respective plurality of levels of goodness, the categorizer system for using one

of the plurality of levels of goodness for invoking an additional categorizer as required; and

a return for returning the item categorized.

24. The system as claimed in claim 23 wherein:

determining the plurality of levels of goodness includes using a process to quantify the plurality of levels of goodness, to prioritize the plurality of levels of goodness, and to resolve two levels of goodness into a third level of goodness.

25. The system as claimed in claim 23 including:

a categorizer system knowledge base for determining the level of goodness for a category with the list of item features.

26. The method as claimed in claim 23 including:

a categorizer system trainer trained using a plurality of items having known categories and the plurality of labels to provide a categorizer system knowledge base.

27. A system for categorization of an item comprising:

a categorizer system knowledge base having a plurality of categories organized in a hierarchy of categories and having respective lists of category features;

a featurizer for featurizing the item to create a list of item features; and

a categorizer system connected to the categorization system knowledge base including:

a plurality of categorizers having one of the plurality of categories, the plurality of categorizers for using the list of item features with the lists of category features to respectively determine a plurality of levels of goodness, the plurality of categorizers categorizing the item in the categorizer system in the plurality of categories based on the respective plurality of levels of goodness,

a mechanism for using one of the plurality of levels of goodness for invoking an additional categorizer of the plurality of categorizers as required; and

a return for returning the item categorized.

28. The system as claimed in claim 27 wherein:

the plurality of categorizers determine the plurality of levels of goodness using a process to quantify the plurality of levels of goodness, to prioritize the

plurality of levels of goodness, and to resolve two levels of goodness into a third level of goodness.

29. The system as claimed in claim 27 wherein:

the categorizer system knowledge base determines the lists of category features.

30. The system as claimed in claim 27 wherein:

the plurality of categorizers include a list mechanism for listing the plurality of categories and the respective levels of goodness; and

the plurality of categorizers categorizes from the list mechanism.

31. The system as claimed in claim 27 wherein:

the return returns one category for the item among the plurality of categories selected from a group consisting of the one category with the best level of goodness for all the plurality of categories and with the best level of goodness for which determining is completed where all of the plurality of categories are not compared.

32. The system as claimed in claim 27 wherein:

the return returns a plurality of categories for the item among the plurality of categories returns a plurality of categories selected from a group consisting of categories up to a fixed number of the plurality of categories, categories having more than a fixed level of goodness, categories fulfilling a user specified preference, categories not from a categorizer, and categories which are a combination thereof.

33. The system as claimed in claim 27 wherein:

the return returns the category for a plurality of items to the categorizer system knowledge base for building a topic hierarchy.

34. The system as claimed in claim 27 including:

a further listing mechanism for listing a plurality of labels for each of the plurality of categories; and

a categorizer system trainer trained using a plurality of items having known categories and the plurality of labels to provide the categorizer system knowledge base.

35. A system for categorization of an item comprising:

a categorizer system knowledge base having a plurality of categories having respective lists of category features;

a featurizer for featurizing the item to create a list of item features; and

a categorizer system connected to the categorizer system knowledge base including:

a plurality of categorizers having the plurality of categories, the plurality of categorizers for determining the list of item features with the lists of category features to respectively determine a plurality of levels of goodness, the plurality of categorizers categorizing the item in the categorizer system in the plurality of categories based on the respective plurality of levels of goodness,

a mechanism for using one of the plurality of levels of goodness for invoking an additional categorizer of the plurality of categorizers as required

a listing mechanism for listing the plurality of categories and the respective levels of goodness on a list, and

a return for returning a category for the item from the list.